Common Transactivation Motif 9aaTAD recruits multiple general co-activators TAF9, MED15, CBP and p300

More than 2000 transcription factors are involved in the human gene regulation to provide cellular function and development. Transactivation domain, TAD, mediates the interaction of transcription factors with general transcriptional co-activators such as TAF9, MED15, CBP and p300. Recently reported NMR-data for interactions of these co-activators with transcription factors reveal tightly binding of KIX and TAZ domains to common 9aaTAD motif. Nine-amino-acid Trans-Activation Domain, 9aaTAD, defines a transactivation domain common to a large super-family of eukaryotic transcription factors represented in yeast by Gal4, Oaf1, Pdr1, Rtg3, Pho4, Gln3, Gcn4 and in mammals by p53, E2A, NFAT, NFkB, HSF1, NF-IL6, MLL, EBNA2, VP16. The 9aaTAD family was derived from the transcription factor Oaf1 and its close orthologs Gal4, Pdr1, Leu3, Tea1 and Cha4.

9aaTAD interactions with a general transcriptional co-activator TAF9/TAFII31 were reported (NMR and biochemical data: p53, VP16, HSF1, NF-IL6, NF-κB and NFAT)[1]. 9aaTAD interactions with KIX domain of further general transcriptional co-activators MED15/Gal11 and CBP-p300 were recently reported (NMR data available for Oaf1, Pdr1 and p53)[2-6]. Oaf1-9aaTAD and Pdr1-9aaTAD[3] upon binding to KIX domain of Med15/Gal11 (9 resp. 12 amino acid long peptides including 9aaTAD) reveal structural similarity, although none of the nine amino acids is conserved in both 9aaTADs (Table 1).

The interaction between TAF9 and p53-TAD1 was restricted by NMR to five amino acids (FSDLW) in the first annotated 9aaTAD of p53 (p53-9aaTAD1). The second annotated 9aaTAD (p53-9aaTAD2) and the pseudo-9aaTAD (p53-9aaTADpse) were located in p53-TAD2 (annotated results from 2007 available on www.expasy.ch/tools). Both p53-9aaTAD1 and p53-9aaTAD2 fulfil the criteria of 9aaTAD pattern, amino acid composition and hydrophobic profile (Fig.1).

Interestingly, CBP-p300 domains KIX, TAZ1, TAZ2 and IBiD are also able to bind p53-9aaTAD1[4,7] (Table 2). The site of interaction between the KIX domain and both p53-TADs corresponds with the annotated 9aaTADs and the TAF9 binding site (NMR data, eight out of nine annotated amino acids). Despite the unrecognizable sequence similarity (Table 2 and 3)[4,7], both p53-TAD1 and p53-TAD2 share the same docking sites on the KIX domain, which may result from the presence of a common motif in the p53 transactivation domains, the 9aaTAD[8]. p53 and further well characterized 9aaTAD-transcription factors, which interact with CBP-p300 and other general co-activator GCN5 are listed in Table 3.

Structural similarities of TADs of p53, STAT1, CITED2 and HIF1a upon binding to TAZ domain were reported by PE Wright lab[9]. Bound TADs are wrapped around TAZ domain in the shared structural order of two helixes with linker (9aaTAD-loop-9aaTAD domain architecture are highlighted in Table 4). Coil-to-helix transition was observed in TAD2 peptide upon its binding to RPA (NMR data)[10] giving disorder-to-order changes into helix-linker-helix conformation (9aaTADpse-loop-9aaTAD2). p53-TAD2 is wrapped around RPA in a very similar way to 9aaTADs of STAT1, CITED2 and HIF1a upon binding to TAZ domain[9] (Table 4). Helix H2 (amino acids 47-55) unlike to helix H1 (amino acids 41-44) has the most extensive buried surface area and therefore appears to be the major determinant of the interaction. Helix H2 is located in the previously annotated p53-9aaTAD2 (amino acids 48-56) and helix H1 in the p53-9aaTADpse (amino acids 38-46) (Table 4). Both helixes are also involved in p53 binding to KIX domain and MDM2[4,7,11].

RPA is the major single-stranded DNA binding protein in the cell. In the recent report[10], Bochkarev and Arrowsmith labs pointed out very impressively that p53-TAD2 imitates the single stranded DNA (peptide bound in the deep basic cleft corresponding to the nucleic acid-binding pocket of the OB fold of RPA). The analogy of 9aaTAD to the DNA implies 9aaTAD in the function of a pseudo DNA-response element recognized by transcription factors of a higher order (9aaTAD mimics the DNA response element in order to recruit general transcription factors). DNA-mimicry for p53 was very recently reported for PC4[12]. Both p53-9aaTAD and VP16-9aaTAD bind to PC4, TAF9, TFB1, MED15/Gal11, CBP-p300, GCN5-PCAF and other cofactors[1,6,13,14].

Diverse transactivation domains[15] were reported to interact with general transcriptional co-activators. None or extremely poor similarities were reported among the transactivation domains in general. Despite the enormous natural diversity, it was possible to generate a pattern and composition rules for 9aaTAD[8]. The establishment of 9aaTAD, its initial pattern and composition definition, identification and annotation (prediction correlated to experimental data) should streamline the investigations in gene regulation by transcription factors and the involvement of multiple transcriptional co-activators. The 9aaTAD prediction and online annotated results are available on www.expasy.ch/tools.

Table 1

Annotated Oaf1/Pip2-9aaTADs:

```
Oaf1 YEAST      D LFDY DFLF
Pip2 YEAST      D FFDY DLLF
Oaf2 CANGA      G VLDF EFLL
Oaf1 ASHGO      D LLDY EFLF
Oaf1 KLULA      D FFDY DFLL
Oaf1 SACMI      D AFWE LLGE
Oaf1 SACPA      T DLEE EYQF
Oaf1 SACBA      T LFDV FIQR
```

Annotated Gal4family-9aaTADs:

```
Gal4 YEAST      D DVYN YLFD
Oaf1 YEAST      D LFDY DFLF
Pdr1 YEAST      E DLYS ILWS
Pdr3 YEAST      G TLDE FVNK
```

Table 2

Interaction	Experiment	p53-TAD Peptides
TAF9 with p53-TAD1	**NMR**	SVEPPLSQ E TFSD LWKL
KO mut. for interaction with TAF9	**NMR**	A A
KIX (CBP) with p53-TAD1	**NMR**	LSQ E TFSD LWKL LPE
KIX (CBP) with p53-TAD2	**NMR**	QAMDDLMLSP D DIEQ WFTE DPGPD
KIX,TAZ1, TAZ2, IBiD (p300)	**NMR**	SQ E TFSD LWKL LPEN
KO mut. with KIX,TAZ1, TAZ2, IBiD	**NMR**	QS
GCN5 with p53-TAD2	**CoIP**	DDLMLSP D DIEQ WFTE
Annotated p53-9aaTAD1:		E TFSD LWKL
Annotated p53-9aaTAD2:		D DIEQ WFTE

Figure 1

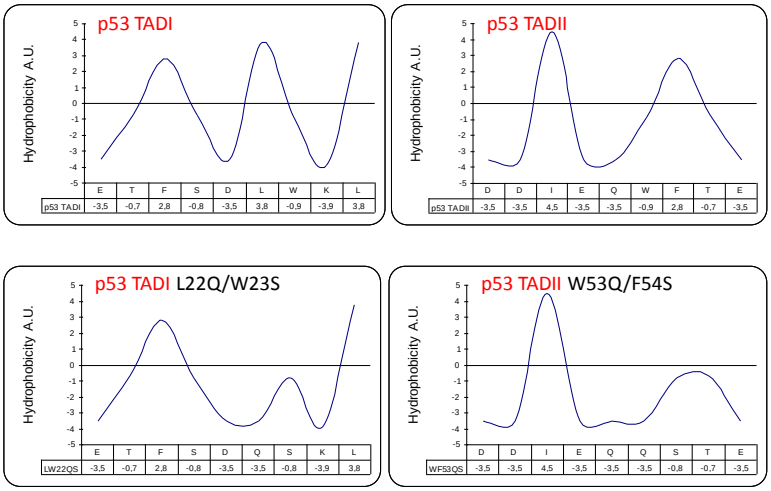

Table 3

Annotated 9aaTADs interacting with multiple co-activators: CBP/p300, TAF9 or GCN5

hp53 TAD1	E TFSD LWKL
cp53 TAD1	E TFSE LWNL
hp53 TAD2	D DIEQ WFTE
cp53 TAD2	E SVVN WLDE
yGal4*	D DVYN YLFD
hNFAT1	D ELDF SILF
pNFAT1	D EFDF SILF
HRX (ALL1/MLL)	S DIMD FVLK
AAC41377.1	S DIME FVLN
MLL2,4	S EIVD FVLK
hSTAT1 TAD1	E EFDE VSRI
hSTAT1 TAD2	V EFDS MMNT
mSTAT1 TAD1	E EFDE MSRI
mSTAT1 TAD2	P EFDS MMST
hSTAT3 TAD1	R TLDS LMQF
hSTAT3 TAD1	E SLTF DMEL
hTF65 TAD1 (NFkB)	S IADM DFSA
hTF65 TAD2	M DFSA LLSQ
HHV1F (VP16)	Q MFTD ALGI
HHV2H	Q MFTD AMGI
hEBNA2	E SWDY IFET
hEBNA2/ W454F	E SFEG IFET
EBNA2 (RHESUS)	E NWDD IFNV
EBNA2 (PAPIO)	N QWED IFNF
hTFE2 (E2A)	D LLDF SMMF
hITF2, hHTF4	D LLDF SAMF
E12 (BRARE)	D LLDF SIMF
yRTG3	E TLDF SLVT
HSF1	S ALLD LFSP
HSF2	D LLVD LFTS
hCEBPA (NF-IL6)	E FLAD LFQH
hCEBPB	D FLSD LFSD
mCEBPB	D FLSD LFAD
hCEBPD	E LFAD LFNS
hCEBPE	Q LLSD LFAV
mCEBPE	Q LLSD LFAM
hKLF2	S VLDF ILSM
mKLF2	N VLDF ILSM
hKLF4	D LLDL DFIL
xKLF4	K FVDL DFIL
E1A_ADECT	D YVLE LLEE
E1A_ADECR	E YVSQ LLED
E1A_ADE41	D DMFQ GLLE
E1A_ADE12	D DILE HLVD
E1A_ADE05	E EMAA SLLD

9aaTADs interacting with CBP-p300 and other co-activators are listed: p53-9aaTADs bind to TAF9[1] and CBP-p300 (KIX, TAZ1, TAZ2 and IBiD domains)[4,7] (Gal4-9aaTAD represent the 9aaTAD family)[8], NFAT1 binds to TAF9 and CBP (TAZ-KIX and TAZ2 domains)[16], MLL/ALL1 binds to TAF9[1] and CBP (KIX and TAZ2 domain)[2,9,17], STAT1 binds to TAZ1 and TAZ2 domain of CBP[9], NFkB binds to TAF9[1] and CBP[1,18], VP16 binds to TAF9, Gal11/MED15, CBP-p300 and GCN5-PCAF[1,6,14], EBNA2 binds to CBP-p300 (N- and C-terminal domains) and PCAF[6], E2A binds to SAGA complex (TAF9-GCN5) and CBP (KIX domain)[19], HSF1 binds to TAF9[1] and CBP[20] and NF-IL6 binds to TAF9[1] and p300[21]. E1A and KLF4 represent CBP-p300 binding transcription factor.

Table 4

	9aaTAD	-	Loop	-	9aaTAD
	E .F.. .F..				E .F.. .F..
CITED2	E EVLM SLVI	G	LDRI		K ELPE LWLG
CITED1	E EVLM SLVV	G	LDRA		N ELPE LWLG
CITED4	E EALT SLEL	G	LHRV		R ELPE LFLG
HIF1a-1	S DLAC RLL	G	QSMDES	GLPQLTSYD	C EVNA PIQG
HIF1a-2	C EVNA PIQ	G	SRNLLQ	G	E ELLR ALDQ
CHICK	C EVNA PIQ	G	NRNLLQ	G	E ELLR ALDQ
HIF2	C EVNV PVL	G	SSTLLQ	G	G DLLR ALDQ
STAT3					
HUMAN	R TLDS LMQF	G	NNGEGAEP SAG	G QF	E SLTF DMEL
XENLA	G TFDS VMQF	PG	EGSE SGN	G NQF	E TLTF DVDL
TETFL	R TLDS LMHN		EAEANP	G HL	D SLTL EMDV
BRARE	R TLDS LMHN		AAEANP	G PL	E SLTL DMEL
STAT1					
HUMAN	E EFDE VSRI V	GS			V EFDS MMNT
MOUSE	E EFDE MSRI V	G			P EFDS MMST
RAT	E EFDE MSKI V	G			S EFDS MMSA
p53-TAD1*	E TFSD LWKL L	P	ENNVLSP	LPS	Q AMDD LMLS
p53-TAD2*	Q AMDD LMLS	P			D DIEQ WFTE
H1 and H2	Q AMDD LMLS	P			D DIEQ WFTE

Disorder-to-order changes giving helix-linker-helix conformation of 9aaTADs are listed (9aaTAD-loop-9aaTAD). CITED2-, HIF1a- and STAT1-TAD share binding regions on TAZ domain of CBP (NMR data[9,22], interacting regions are in bold and underlined). p53-TAD2 binds to RPA[10] in a similar manner to the TADs of STAT1, CITED2 and HIF1a to TAZ. *Sequence similarity of STAT1- and p53-TAD was previously reported (NMR data, amino acids interacting with CBP are in bold and underlined)[9].

1 M. Uesugi, O. Nyanguile, H. Lu et al., *Science* **277** (5330), 1310 (1997); M. Uesugi and G. L. Verdine, *Proc Natl Acad Sci U S A* **96** (26), 14801 (1999); Y. Choi, S. Asada, and M. Uesugi, *J Biol Chem* **275** (21), 15912 (2000).

2 C. W. Lee, M. Arai, M. A. Martinez-Yamout et al., *Biochemistry* (2009); N. K. Goto, T. Zor, M. Martinez-Yamout et al., *J Biol Chem* **277** (45), 43168 (2002); I. Radhakrishnan, G. C. Perez-Alvarado, D. Parker et al., *Cell* **91** (6), 741 (1997); T. Zor, B. M. Mayr, H. J. Dyson et al., *J Biol Chem* **277** (44), 42241 (2002).

3 J. K. Thakur, H. Arthanari, F. Yang et al., *Nature* **452** (7187), 604 (2008); J. K. Thakur, H. Arthanari, F. Yang et al., *J Biol Chem* **284** (7), 4422 (2009).

4 A. M. Gamper and R. G. Roeder, *Mol Cell Biol* **28** (8), 2517 (2008).

5 M. Shuen, N. Avvakumov, P. G. Walfish et al., *J Biol Chem* **277** (34), 30844 (2002).

6 L. Wang, S. R. Grossman, and E. Kieff, *Proc Natl Acad Sci U S A* **97** (1), 430 (2000).

7 D. P. Teufel, S. M. Freund, M. Bycroft et al., *Proc Natl Acad Sci U S A* **104** (17), 7009 (2007); D. P. Teufel, M. Bycroft, and A. R. Fersht, *Oncogene* **28** (20), 2112 (2009); H. Feng, L. M. Jenkins, S. R. Durell et al., *Structure* **17** (2), 202 (2009); L. M. Jenkins, H. Yamaguchi, R. Hayashi et al., *Biochemistry* (2009); J. C. Ferreon, C. W. Lee, M. Arai et al., *Proc Natl Acad Sci U S A* **106** (16), 6591 (2009).

8 S. Piskacek, M. Gregor, M. Nemethova et al., *Genomics* **89** (6), 756 (2007).

9 J. M. Wojciak, M. A. Martinez-Yamout, H. J. Dyson et al., *Embo J* **28** (7), 948 (2009).

10 L. Kaustov, G. S. Yi, A. Ayed et al., *Cell cycle (Georgetown, Tex* **5** (5), 489 (2006).

11 S. W. Chi, S. H. Lee, D. H. Kim et al., *J Biol Chem* **280** (46), 38795 (2005).

12 S. Rajagopalan, A. Andreeva, D. P. Teufel et al., *J Biol Chem* (2009).

13 C. Langlois, C. Mas, P. Di Lello et al., *Journal of the American Chemical Society* **130** (32), 10596 (2008).

14 T. Fukasawa, M. Fukuma, K. Yano et al., *DNA Res* **8** (1), 23 (2001); L. Badi and A. Barberis, *Mol Genet Genomics* **265** (6), 1076 (2001); Y. J. Kim, S. Bjorklund, Y. Li et al., *Cell* **77** (4), 599 (1994); Y. Suzuki, Y. Nogi, A. Abe et al., *Mol Cell Biol* **8** (11), 4991 (1988); J. S. Fassler and F. Winston, *Mol Cell Biol* **9** (12), 5602 (1989); J. M. Park, H. S. Kim, S. J. Han et al., *Mol Cell Biol* **20** (23), 8709 (2000); Z. Lu, A. Z. Ansari, X. Lu et al., *Proc Natl Acad Sci U S A* **99** (13), 8591 (2002); M. J. Swanson, H. Qiu, L. Sumibcay et al., *Mol Cell Biol* **23** (8), 2800 (2003); G. O. Bryant and M. Ptashne, *Mol Cell* **11** (5), 1301 (2003); J. Fishburn, N. Mohibullah, and S. Hahn, *Mol Cell* **18** (3), 369 (2005); M. K. Lim, V. Tang, A. Le Saux et al., *J Mol Biol* **374** (1), 9 (2007); S. Lallet, H. Garreau, C. Garmendia-Torres et al., *Molecular microbiology* **62** (2), 438 (2006); M. Dietz, W. T. Heyken, J. Hoppen et al., *Molecular microbiology* **48** (4), 1119 (2003); T. Mizuno and S. Harashima, *Mol Genet Genomics* **269** (1), 68 (2003); C. Rachez, B. D. Lemon, Z. Suldan et al., *Nature* **398** (6730), 824 (1999); K. Ikeda, T. Stuehler, and M. Meisterernst, *Genes Cells* **7** (1), 49 (2002); T. Uhlmann, S. Boeing, M. Lehmbacher et al., *J Biol Chem* **282** (4), 2163 (2007).

15 A. H. Brivanlou and J. E. Darnell, Jr., *Science* **295** (5556), 813 (2002).

16 C. Garcia-Rodriguez and A. Rao, *The Journal of experimental medicine* **187** (12), 2031 (1998).

17 R. Prasad, T. Yano, C. Sorio et al., *Proc Natl Acad Sci U S A* **92** (26), 12160 (1995); P. Ernst, J. Wang, M. Huang et al., *Mol Cell Biol* **21** (7), 2249 (2001); S. Bruschweiler, P. Schanda, K. Kloiber et al., *Journal of the American Chemical Society* (2009).

18 G. H. Liu, J. Qu, and X. Shen, *Biochimica et biophysica acta* **1783** (5), 713 (2008).

19 M. E. Massari, P. A. Grant, M. G. Pray-Grant et al., *Mol Cell* **4** (1), 63 (1999); R. Bayly, T. Murase, B. D. Hyndman et al., *Mol Cell Biol* **26** (17), 6442 (2006).

20 S. Hong, S. H. Kim, M. A. Heo et al., *FEBS letters* **559** (1-3), 165 (2004).

21 S. Mink, B. Haenig, and K. H. Klempnauer, *Mol Cell Biol* **17** (11), 6609 (1997).

22 S. A. Dames, M. Martinez-Yamout, R. N. De Guzman et al., *Proc Natl Acad Sci U S A* **99** (8), 5271 (2002).